

## **STIC Biotechnology Systems Branch**

### **RAW SEQUENCE LISTING** **ERROR REPORT**

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 10/524,053  
Source: PG 1/10  
Date Processed by STIC: 2/28/05

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION AND PATENTIN SOFTWARE QUESTIONS, PLEASE CONTACT MARK SPENCER, TELEPHONE: 571-272-2510; FAX: 571-273-0221

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE **CHECKER VERSION 4.2.2 PROGRAM**, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW FOR ADDRESS:

<http://www.uspto.gov/web/offices/pac/checker/chkrnote.htm>

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail. Please consider using alternate methods of submission for the disk/CD-Rom or replacement disk/CD-Rom. Any reply including a sequence listing in electronic form should NOT be sent to the 20231 zip code address for the United States Patent and Trademark Office, and instead should be sent via the following to the indicated addresses:

1. EFS-Bio (<<http://www.uspto.gov/ebc/efs/downloads/documents.htm>> , EFS Submission User Manual - ePAVE)
2. U.S. Postal Service: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450
3. Hand Carry, Federal Express, United Parcel Service, or other delivery service (EFFECTIVE 01/14/05):  
U.S. Patent and Trademark Office, Mail Stop Sequence, Customer Window, Randolph Building, 401 Dulany Street, Alexandria, VA 22314

Revised 01/24/05

## Raw Sequence Listing Error Summary

### ERROR DETECTED

### SUGGESTED CORRECTION

SERIAL NUMBER: 10/524,053

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1      Wrapped Nucleics  
    Wrapped Aminos      The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."
- 2      Invalid Line Length      The rules require that a line not exceed 72 characters in length. This includes white spaces.
- 3      Misaligned Amino  
    Numbering      The numbering under each 5<sup>th</sup> amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.
- 4      Non-ASCII      The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.
- 5      Variable Length      Sequence(s)      contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.
- 6      PatentIn 2.0  
    "bug"      A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s)     . Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.
- 7      Skipped Sequences  
    (OLD RULES)      Sequence(s)      missing. If intentional, please insert the following lines for each skipped sequence:  
                                 (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
                                 (i)      SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading)  
                                 (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
                                 This sequence is intentionally skipped  
  
                                 Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.
- 8      Skipped Sequences  
    (NEW RULES)      Sequence(s) 11 missing. If intentional, please insert the following lines for each skipped sequence.  
                                 <210> sequence id number  
                                 <400> sequence id number  
                                 000
- 9      Use of n's or Xaa's  
    (NEW RULES)      Use of n's and/or Xaa's have been detected in the Sequence Listing.  
                                 Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present.  
                                 In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.
- 10      Invalid <213>  
    Response      Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence
- 11      Use of <220>      Sequence(s)      missing the <220> "Feature" and associated numeric identifiers and responses  
                                 Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section.  
                                 (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)
- 12      PatentIn 2.0  
    "bug"      Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.
- 13      Misuse of n/Xaa      "n" can only represent a single nucleotide; "Xaa" can only represent a single amino acid



PCT

## RAW SEQUENCE LISTING

DATE: 02/28/2005

PATENT APPLICATION: US/10/524,053

TIME: 10:13:31

Input Set : A:\DOCSMTL-#1207587-v1-SEQLIST.TXT

Output Set: N:\CRF4\02282005\J524053.raw

4 <110> APPLICANT: NATIONAL RESEARCH COUNCIL OF CANADA  
 5 SU, Zhengding  
 6 NI, Feng  
 8 <120> TITLE OF INVENTION: Novel Fusion Proteins for the Production  
 9 of Recombinant Peptides  
 12 <130> FILE REFERENCE: 2139  
 C--> 14 <140> CURRENT APPLICATION NUMBER: US/10/524,053  
 C--> 14 <141> CURRENT FILING DATE: 2005-02-08  
 14 <150> PRIOR APPLICATION NUMBER: US 60/402,075  
 15 <151> PRIOR FILING DATE: 2002-08-09  
 17 <160> NUMBER OF SEQ ID NOS: 48  
 19 <170> SOFTWARE: FastSEQ for Windows Version 4.0  
 21 <210> SEQ ID NO: 1  
 22 <211> LENGTH: 25  
 23 <212> TYPE: PRT  
 24 <213> ORGANISM: Staphylococcus nuclease  
 26 <400> SEQUENCE: 1  
 27 Ala Thr Ser Thr Lys Lys Leu His Lys Glu Pro Ala Thr Leu Ile Lys  
 28 1 5 10 15  
 29 Ala Ile Asp Gly Asp Thr Val Lys Leu  
 30 20 25  
 32 <210> SEQ ID NO: 2  
 33 <211> LENGTH: 5  
 34 <212> TYPE: PRT  
 35 <213> ORGANISM: Staphylococcus nuclease  
 37 <400> SEQUENCE: 2  
 38 Tyr Lys Gly Gln Pro  
 39 1 5  
 41 <210> SEQ ID NO: 3  
 42 <211> LENGTH: 29  
 43 <212> TYPE: PRT  
 44 <213> ORGANISM: Staphylococcus nuclease  
 46 <400> SEQUENCE: 3  
 47 Leu Leu Leu Val Asp Thr Pro Glu Thr Lys His Pro Lys Lys Gly Val  
 48 1 5 10 15  
 49 Glu Lys Tyr Gly Pro Glu Ala Ser Ala Phe Thr Lys Lys  
 50 20 25  
 52 <210> SEQ ID NO: 4  
 53 <211> LENGTH: 32  
 54 <212> TYPE: PRT  
 55 <213> ORGANISM: Staphylococcus nuclease  
 57 <400> SEQUENCE: 4  
 58 Val Glu Asn Ala Lys Lys Ile Glu Val Glu Phe Asp Lys Gly Gln Arg

Does Not Comply  
Corrected Diskette Needed

pp 2-3, 5-6

## RAW SEQUENCE LISTING

DATE: 02/28/2005

PATENT APPLICATION: US/10/524,053

TIME: 10:13:31

Input Set : A:\DOCSMTL-#1207587-v1-SEQLIST.TXT

Output Set: N:\CRF4\02282005\J524053.raw

59 1 5 10 15  
60 Thr Asp Lys Tyr Gly Arg Gly Leu Ala Tyr Ile Tyr Ala Asp Gly Lys  
61 20 25 30  
63 <210> SEQ ID NO: 5  
64 <211> LENGTH: 5  
65 <212> TYPE: PRT  
66 <213> ORGANISM: Staphylococcus nuclease  
68 <400> SEQUENCE: 5  
69 Val Asn Glu Ala Leu  
70 1 5  
72 <210> SEQ ID NO: 6  
73 <211> LENGTH: 6  
74 <212> TYPE: PRT  
75 <213> ORGANISM: Staphylococcus nuclease  
77 <400> SEQUENCE: 6  
78 Phe Asn Pro Arg Gly Ser  
79 1 5  
81 <210> SEQ ID NO: 7  
82 <211> LENGTH: 14  
83 <212> TYPE: PRT  
84 <213> ORGANISM: Staphylococcus nuclease  
86 <400> SEQUENCE: 7  
87 Val Arg Gln Gly Leu Ala Lys Val Ala Tyr Val Tyr Lys Pro  
88 1 5 10  
90 <210> SEQ ID NO: 8  
91 <211> LENGTH: 32  
92 <212> TYPE: PRT  
93 <213> ORGANISM: Staphylococcus nuclease  
95 <400> SEQUENCE: 8  
96 Asn Asn Thr His Glu Gln Leu Leu Arg Lys Ser Glu Ala Gln Ala Lys  
97 1 5 10 15  
98 Lys Glu Lys Leu Asn Ile Trp Ser Glu Asp Asn Ala Asp Ser Gly Gln  
99 20 25 30  
101 <210> SEQ ID NO: 9  
102 <211> LENGTH: 4  
103 <212> TYPE: PRT  
104 <213> ORGANISM: Artificial Sequence  
106 <220> FEATURE:  
107 <223> OTHER INFORMATION: Peptidic Cleavage Site  
109 <400> SEQUENCE: 9  
110 Phe Asn Pro Arg  
111 1  
113 <210> SEQ ID NO: 10  
114 <211> LENGTH: 4  
115 <212> TYPE: PRT  
116 <213> ORGANISM: Artificial Sequence  
118 <220> FEATURE:  
119 <223> OTHER INFORMATION: Peptidic Cleavage Site  
121 <400> SEQUENCE: 10

*what is the source  
of the genetic material?  
(see item 11 on  
Error summary  
sheet)*

## RAW SEQUENCE LISTING

DATE: 02/28/2005

PATENT APPLICATION: US/10/524,053

TIME: 10:13:31

Input Set : A:\DOCSMTL-#1207587-v1-SEQLIST.TXT

Output Set: N:\CRF4\02282005\J524053.raw

122 Leu Val Pro Arg  
 123 1  
 125 <210> SEQ ID NO: 11  
 127 <220> FEATURE:  
 128 <223> OTHER INFORMATION: Peptidic Cleavage Site  
 130 <400> SEQUENCE: 11  
 W--> 131 000  
 133 <210> SEQ ID NO: 12  
 134 <211> LENGTH: 5  
 135 <212> TYPE: PRT  
 136 <213> ORGANISM: Artificial Sequence  
 138 <220> FEATURE:  
 139 <223> OTHER INFORMATION: Peptidic Cleavage Site  
 141 <400> SEQUENCE: 12  
 142 Asp Asp Asp Asp Lys  
 143 1 5  
 145 <210> SEQ ID NO: 13  
 146 <211> LENGTH: 351  
 147 <212> TYPE: DNA  
 148 <213> ORGANISM: Artificial Sequence  
 150 <220> FEATURE:  
 151 <223> OTHER INFORMATION: Fusion Carrier Protein named SFC120  
 153 <220> FEATURE:  
 154 <221> NAME/KEY: CDS  
 155 <222> LOCATION: (1)...(351)  
 157 <400> SEQUENCE: 13  
 158 gca act tca act aaa aaa tta cat aaa gaa cct gcg act tta att aaa 48  
 159 Ala Thr Ser Thr Lys Lys Leu His Lys Glu Pro Ala Thr Leu Ile Lys  
 160 1 5 10 15  
 162 gcg att gat ggt gat acg gtt aaa tta atg tac aaa ggt caa cca atg 96  
 163 Ala Ile Asp Gly Asp Thr Val Lys Leu Met Tyr Lys Gly Gln Pro Met  
 164 20 25 30  
 166 aca ttc aga cta tta ttg gtt gat aca cct gaa aca aag cat cct aaa 144  
 167 Thr Phe Arg Leu Leu Leu Val Asp Thr Pro Glu Thr Lys His Pro Lys  
 168 35 40 45  
 170 aaa ggt gta gag aaa tat ggt cct gaa gca agt gca ttt acg aaa aaa 192  
 171 Lys Gly Val Glu Lys Tyr Gly Pro Glu Ala Ser Ala Phe Thr Lys Lys  
 172 50 55 60  
 174 atg gta gaa aat gca aag aaa att gaa gtc gag ttt gac aaa ggt caa 240  
 175 Met Val Glu Asn Ala Lys Lys Ile Glu Val Glu Phe Asp Lys Gly Gln  
 176 65 70 75 80  
 178 aga act gat aaa tat gga cgt ggc tta gcg tat att tat gct gat gga 288  
 179 Arg Thr Asp Lys Tyr Gly Arg Gly Leu Ala Tyr Ile Tyr Ala Asp Gly  
 180 85 90 95  
 182 aaa atg gta aac gaa gct tta gtt cgt caa ggc ttg gct aaa gtt gct 336  
 183 Lys Met Val Asn Glu Ala Leu Val Arg Gln Gly Leu Ala Lys Val Ala  
 184 100 105 110  
 186 tat gtt tac aaa cct 351  
 187 Tyr Val Tyr Lys Pro

*delete this section - see item 8  
 on Error  
 summary sheet  
 for correct format*

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/524,053

DATE: 02/28/2005

TIME: 10:13:31

Input Set : A:\DOCSMTL-#1207587-v1-SEQLIST.TXT

Output Set: N:\CRF4\02282005\J524053.raw

```

188          115
191 <210> SEQ ID NO: 14
192 <211> LENGTH: 117
193 <212> TYPE: PRT
194 <213> ORGANISM: Artificial Sequence
196 <220> FEATURE:
197 <223> OTHER INFORMATION: Fusion Carrier Protein named SFC120
199 <400> SEQUENCE: 14
200 Ala Thr Ser Thr Lys Lys Leu His Lys Glu Pro Ala Thr Leu Ile Lys
201   1          5          10          15
202 Ala Ile Asp Gly Asp Thr Val Lys Leu Met Tyr Lys Gly Gln Pro Met
203          20          25          30
204 Thr Phe Arg Leu Leu Leu Val Asp Thr Pro Glu Thr Lys His Pro Lys
205          35          40          45
206 Lys Gly Val Glu Lys Tyr Gly Pro Glu Ala Ser Ala Phe Thr Lys Lys
207   50          55          60
208 Met Val Glu Asn Ala Lys Lys Ile Glu Val Glu Phe Asp Lys Gly Gln
209 65          70          75          80
210 Arg Thr Asp Lys Tyr Gly Arg Gly Leu Ala Tyr Ile Tyr Ala Asp Gly
211          85          90          95
212 Lys Met Val Asn Glu Ala Leu Val Arg Gln Gly Leu Ala Lys Val Ala
213          100          105          110
214 Tyr Val Tyr Lys Pro
215          115
217 <210> SEQ ID NO: 15
218 <211> LENGTH: 369
219 <212> TYPE: DNA
220 <213> ORGANISM: Artificial Sequence
222 <220> FEATURE:
223 <223> OTHER INFORMATION: Fusion Carrier Protein named SFC120 ( N118E,
224     N119F)
226 <220> FEATURE:
227 <221> NAME/KEY: CDS
228 <222> LOCATION: (1)...(369)
230 <400> SEQUENCE: 15
231 aca gcc atg gca act tca act aaa aaa tta cat aaa gaa cct gcg act      48
232 Thr Ala Met Ala Thr Ser Thr Lys Lys Leu His Lys Glu Pro Ala Thr
233   1          5          10          15
235 tta att aaa gcg att gat ggt gat acg gtt aaa tta atg tac aaa ggt      96
236 Leu Ile Lys Ala Ile Asp Gly Asp Thr Val Lys Leu Met Tyr Lys Gly
237          20          25          30
239 caa cca atg aca ttc aga cta tta ttg gtt gat aca cct gaa aca aag      144
240 Gln Pro Met Thr Phe Arg Leu Leu Leu Val Asp Thr Pro Glu Thr Lys
241          35          40          45
243 cat cct aaa aaa ggt gta gag aaa tat ggt cct gaa gca agt gca ttt      192
244 His Pro Lys Lys Gly Val Glu Lys Tyr Gly Pro Glu Ala Ser Ala Phe
245          50          55          60
247 acg aaa aaa atg gta gaa aat gca aag aaa att gaa gtc gag ttt gac      240
248 Thr Lys Lys Met Val Glu Asn Ala Lys Lys Ile Glu Val Glu Phe Asp

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/524,053

DATE: 02/28/2005

TIME: 10:13:31

Input Set : A:\DOCSMTL-#1207587-v1-SEQLIST.TXT

Output Set: N:\CRF4\02282005\J524053.raw

```

249 65          70          75          80
251 aaa ggt caa aga act gat aaa tat gga cgt ggc tta gcg tat att tat      288
252 Lys Gly Gln Arg Thr Asp Lys Tyr Gly Arg Gly Leu Ala Tyr Ile Tyr
253          85          90          95
255 gct gat gga aaa atg gta aac gaa gct tta gtt cgt caa ggc ttg gct      336
256 Ala Asp Gly Lys Met Val Asn Glu Ala Leu Val Arg Gln Gly Leu Ala
257          100          105          110
259 aaa gtt gct tat gtt tac aaa cct gaa ttc atg      369
260 Lys Val Ala Tyr Val Tyr Lys Pro Glu Phe Met
261          115          120
264 <210> SEQ ID NO: 16
265 <211> LENGTH: 123
266 <212> TYPE: PRT
267 <213> ORGANISM: Artificial Sequence
269 <220> FEATURE:
270 <223> OTHER INFORMATION: Fusion Carrier Protein named SFC120 ( N118E,
271 N119F)
273 <400> SEQUENCE: 16
274 Thr Ala Met Ala Thr Ser Thr Lys Lys Leu His Lys Glu Pro Ala Thr
275 1          5          10          15
276 Leu Ile Lys Ala Ile Asp Gly Asp Thr Val Lys Leu Met Tyr Lys Gly
277          20          25          30
278 Gln Pro Met Thr Phe Arg Leu Leu Leu Val Asp Thr Pro Glu Thr Lys
279          35          40          45
280 His Pro Lys Lys Gly Val Glu Lys Tyr Gly Pro Glu Ala Ser Ala Phe
281          50          55          60
282 Thr Lys Lys Met Val Glu Asn Ala Lys Lys Ile Glu Val Glu Phe Asp
283 65          70          75          80
284 Lys Gly Gln Arg Thr Asp Lys Tyr Gly Arg Gly Leu Ala Tyr Ile Tyr
285          85          90          95
286 Ala Asp Gly Lys Met Val Asn Glu Ala Leu Val Arg Gln Gly Leu Ala
287          100          105          110
288 Lys Val Ala Tyr Val Tyr Lys Pro Glu Phe Met
289          115          120
291 <210> SEQ ID NO: 17
292 <211> LENGTH: 393
293 <212> TYPE: DNA
294 <213> ORGANISM: Artificial Sequence
296 <220> FEATURE:
297 <223> OTHER INFORMATION: pHSN-Met65Leu
299 <220> FEATURE:
300 <221> NAME/KEY: CDS
301 <222> LOCATION: (1)...(393)
303 <400> SEQUENCE: 17
304 aca gcc atg gga cac cat cac cat cac cat ggc gca act tca act aaa
305 Thr Ala Met Gly His His His His His His Gly Ala Thr Ser Thr Lys
306 1          5          10          15
308 aaa tta cat aaa gaa cct gcg act tta att aaa gcg att gat ggt gat
309 Lys Leu His Lys Glu Pro Ala Thr Leu Ile Lys Ala Ile Asp Gly Asp

```

*what is the source of genetic material?*

*Please correct this type of error in subsequent sequences*

## RAW SEQUENCE LISTING ERROR SUMMARY

DATE: 02/28/2005

PATENT APPLICATION: US/10/524,053

TIME: 10:13:32

FYI

Input Set : A:\DOCSMTL-#1207587-v1-SEQLIST.TXT

Output Set: N:\CRF4\02282005\J524053.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:41; Xaa Pos. 20,21,27

Seq#:45; Xaa Pos. 22

Seq#:48; Xaa Pos. 3,5,6,8,10,11



## VERIFICATION SUMMARY

DATE: 02/28/2005

PATENT APPLICATION: US/10/524,053

TIME: 10:13:32

Input Set : A:\DOCSMTL-#1207587-v1-SEQLIST.TXT

Output Set: N:\CRF4\02282005\J524053.raw

L:14 M:270 C: Current Application Number differs, Replaced Current Application No  
L:14 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
L:131 M:300 W: (50) Intentionally skipped Sequence, : Sequence Id (11) SEQUENCE:  
L:767 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:41 after pos.:16  
L:817 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:45  
L:824 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:45 after pos.:16  
L:891 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:48 after pos.:0